Thema
UMR 8184
THéorie Économique, Modélisation et Applications

THEMA Working Paper n°2017-12
Université de Cergy-Pontoise, France

Sibling correlations in terms of education, profession and earnings, in France

Céline Lecavelier des Etangs-Levallois, Arnaud Lefranc

UNIVERSITÉ de Cergy-Pontoise
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

April 2017

Sibling correlations in terms of education, profession and earnings, in France[1]

Céline Lecavelier des Etangs-Levallois[2], Arnaud Lefranc

Abstract

This paper examines the contribution of family background to inequality in France by estimating sibling correlations in various measures of socio-economic success. Compared to often reported measures of intergenerational elasticity, the sibling correlation in socio-economic outcomes allows to capture a broader set of family influences. We use data from the French Education-Training-Employment (FQP) survey to investigate similarities between siblings in education, social prestige and earnings. We also investigate trends over time in sibling correlations and differences across family types in siblings' characteristics. Our results indicate a high degree of association in siblings' socio-economic success. The correlation is around 0.3 and 0.5 respectively for social prestige and years of education. The sibling correlation in annual earnings is around 0.4. All in all, this indicates that estimates of the intergenerational elasticity lead to underestimate the role of family background in children's success in France.

Keywords: sibling correlations, intergenerational mobility, family background

JEL classification: D31, J62

---

[1]We thank the participants of the EALE conference 2016 in Ghent (Belgium), the AFSE conference 2016 in Nancy (France), the JMA conference 2016 in Besançon (France) and the Labex MME-DIIs internal Seminar in Cergy (France), for their helpful comments. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

[2]THEMA - Université de Cergy-Pontoise, 33 boulevard du Port - 95011 Cergy-Pontoise cedex (France); celine.lecavelier[at]u-cergy.fr

1 Introduction

Recent public debates have echoed growing concern that, in modern democratic societies, a sizable share of economic inequality remains inherited within families. Assessing the role of family background in shaping individual economic success is indeed crucial to gauge the degree of inequality of opportunity in a society. Over the last twenty years, an important body of empirical research has investigated the extent of the intergenerational transmission of inequality (Solon, 1999; Björklund and Jäntti, 2009; Black and Devereux, 2011). This literature has demonstrated that between 20 and 60% of economic advantage is transmitted, within families, from one generation to the next. In the case of France, existing estimates indicate a value of the intergenerational earnings elasticity of about 0.5 (Lefranc, 2011). This suggests that about 25% of earnings inequality is transmitted within families across generations. However, by focusing on a single dimension of family characteristics, namely earnings, estimates of the intergenerational earnings elasticity fall short of fully accounting for the variety of channels through which the characteristics of the family influence the outcomes of their offspring. To provide a more comprehensive view, several recent papers have examined the degree of association in siblings' social and economic success, as surveyed in Björklund and Jäntti (2009, 2012) and Björklund and Salvanes (2010). Compared to the intergenerational correlation, the sibling correlation in socioeconomic outcomes allows to capture a broader set of family influences.

Sibling analysis is also employed in sociological studies to assess more than the impact of parental education or occupation, as in Boutchénik et al. (2015), Sieben (2001) or Knigge (2015). Sibling correlations indeed provide a summary measure of all the effects attributed to factors shared by siblings (Björklund and Jäntti, 2009; Björklund et al., 2010). It captures the overall impact of growing up in the same family, and thus allows a more complete investigation of the role of family background in inequalities. As explained in Solon (1999), Björklund and Jäntti (2009) and Bingley and Cappellari (2013), the sibling association in one particular measure of socioeconomic success (e.g. earnings, SES, education, etc.) can be decomposed as the sum of two terms. The first one is the square of the intergenerational correlation in the specific measure of socioeconomic success (e.g. the intergenerational correlation in earnings). The second one captures the influence of all the factors shared among siblings that are uncorrelated with the relevant parental measure of socioeconomic success.

Sibling correlations for various socioeconomic outcomes have been estimated in different countries, since early studies focusing on the US as Corcoran et al. (1976) or later Altonji and Dunn (1991), Corcoran et al. (1990) and Solon et al. (1991). In terms of education, sibling correlations are recently found to lie between 0.4 for Nordic countries (Björklund et al., 2009; Björklund and Salvanes, 2010; Björklund and Jäntti, 2012) and 0.6 for the United States (Conley and Glauber, 2008; Mazumder, 2008, 2011). Correlations in earnings lie between 0.2 for Nordic countries (Björklund et al., 2002; Björklund and Jäntti, 2009, 2012; Schnitzlein, 2014) and 0.4 for Germany (Schnitzlein, 2014) and the United States (Björklund et al., 2002; Conley and Glauber, 2008; Levine and Mazumder, 2007; Mazumder, 2008, 2011; Schnitzlein, 2014). Thus these sibling correlation estimates reveal a far more important transmission of inequalities than shown by estimated intergenerational elasticities. Nevertheless, if several authors have investigated the cases of other countries, the extent of sibling correlation in economic success has not been investigated for France. Boutchénik et al. (2015) study siblings' resemblance in terms of education and profession but not in terms of earnings. Our objective is to fill this gap in the literature.

Several features of France's socioeconomic setting make this country an interesting case for studying the role of family influences in shaping inequality of opportunity. Firstly, the French labor market is largely viewed as a heavily regulated one yielding a more compressed wage structure than observed in Anglo-Saxon economies.

Secondly, deep reforms of the educational system have led to an important rise in access to higher education[3]. Furthermore, it is worth recalling that college, university and "grandes écoles" education are free of tuition in France. However, in international comparisons, France stands out as a country with low intergenerational mobility and high inequality of opportunity. In this context, it is worth investigating further the role of family background by relying on more comprehensive measure of family influences.

The objective of this paper is to measure the extent of sibling association in socioeconomic outcomes in France. We use data from the French Education-Training-Employment (FQP) survey to estimate sibling similarities in different socioeconomic outcomes in France: profession, education and earnings. We find sibling correlations around 0.3, 0.4 and 0.5 for prestige scores, annual earnings and education years respectively. When conducting a study by gender, it appears that same-sex siblings have more in common than in mixed pairs, for each outcome. Additional parameters are then investigated. They do not lead to any clear conclusion toward the evolution in time of sibling correlations. However concerning the impact of family composition, closely spaced siblings are more alike and family size seems to have a positive effect on sibling correlations. Finally we investigate the effect of parental education and profession but observe no clear pattern, except for the decrease of sibling correlations in earnings with higher educational levels of both parents.

The rest of the paper is organized as follows. Section 2 presents the estimation methods of sibling correlations. Section 3 presents the FQP data and describes the construction of the outcomes we further investigate. Section 4 reports the results obtained for France. Section 5 concludes.

[3]Before 1975, lower secondary education was segmented into vocational and general schooling. This dual system was reformed in 1975 under the "réforme Haby" to create a unified junior high school curriculum. Access to higher education rose markedly, first in the late sixties-early seventies and again, during the late eighties-early nineties.

2 Methods

2.1 Pearson's and polychoric correlations

In order to estimate linear sibling correlation coefficients, let y_{ijt} be a continuous outcome of individual j in family i at time t. The importance of family background is measured by the share of variance of the long-run outcome that is accounted for by family effects. This "R^2" of family background is called the sibling correlation since it coincides with the correlation coefficient of randomly drawn pairs of siblings.

The modeled outcome y_{ijt} consists of a permanent part y_{ij} decomposed in a family component a_i, common to both siblings in family i and an individual-specific component b_{ij} for sibling j in family i, as well as a transitory error v_{ijt}:

$$y_{ijt} = y_{ij} + v_{ijt} = a_i + b_{ij} + v_{ijt}, \quad a \perp b \perp v. \tag{1}$$

The variance of the long-run outcome equals the sum of the variances of the family and individual components:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2. \tag{2}$$

Thus the share of family background in the long-run outcome variance, the sibling correlation in permanent outcome ρ, is:

$$\rho = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{3}$$

A set of complementary controls can be included in the estimation of the model in order to first purge the long-run outcome of some effects:

$$y_{ij} = \gamma Z_{ij} + a'_i + b'_{ij} = \gamma Z_{ij} + e_{ij}. \tag{4}$$

In particular the vector Z_{ij} contains in our French estimations a gender dummy, a quartic function of age, and all corresponding interactions. The residuals e_{ij} from the regression equation, free of gender and age effects, are then used in order to compute Pearson's correlations between two siblings 1 and 2:

$$\rho_{e_1,e_2} = \frac{cov(e_1, e_2)}{\sigma_{e_1}\sigma_{e_2}}. \tag{5}$$

In addition to linear sibling correlations, we also estimate polychoric correlations. Based on the same model, they measure the association between two ordinal variables assumed to be determined by a latent continuous variable following a bivariate normal distribution (Drasgow, 1988).

2.2 Sibling correlations linked to intergenerational elasticities

To link the sibling approach to the estimation of the transmission of socioeconomic outcomes using the intergenerational elasticity (IGE) β, we can further decompose our family component a_i into a part correlated to father's outcome X_i and all other family factors v_i uncorrelated with it:

$$a_i = \beta X_i + v_i. \tag{6}$$

Then son's outcome regressed on father's one is now:

$$y_{ij} = \beta X_i + u_{ij}, \text{ with } u_{ij} = v_i + b_{ij}. \tag{7}$$

Expressing the variance of the family component a_i yields:

$$\sigma_a^2 = \beta^2 \sigma_X^2 + \sigma_v^2, \tag{8}$$

and dividing both sides by σ_y^2, assuming same distributions among both generations, i.e. $\sigma_y = \sigma_X$, gives the sibling correlation ρ:

$$\rho = \frac{\sigma_a^2}{\sigma_y^2} = \beta^2 + \frac{\sigma_v^2}{\sigma_y^2}. \tag{9}$$

The sibling correlation ρ can thus be expressed as the sum of two terms: the squared of the IGE[4] β and a second component capturing the effect of all factors shared among siblings and uncorrelated with father's outcome. We can use this decomposition to consider how much of the sibling correlation is related to each part and thus how much is unaccounted for by the IGE.

2.3 Correlations on predicted variables

As further explained in Section 3, for the case of France, we do not estimate correlations on directly observed variables. Indeed extensive information is available for one of the siblings – hereafter "ego" – but only limited information is available for the other one – hereafter "alter". Instead we first predict continuous variables to then use them to investigate sibling correlations. We can model the latent outcome y as the sum of our predicted variable $\hat{y}$ and an ϵ, for each sibling:

$$y_j = \hat{y}_j + \epsilon_j, \text{with } j = E \text{ for ego, } j = A \text{ for alter.} \tag{10}$$

Considering that the distributions are the same for both siblings (that is $\sigma_{\hat{y_E}} = \sigma_{\hat{y_A}}$ and $\sigma_{\epsilon_E} = \sigma_{\epsilon_A}$) and that $\hat{y}$ and ϵ are independent (so $\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_\epsilon^2$), we find:

$$\rho(y_E, y_A) = \frac{cov(\hat{y}_E, \hat{y}_A) + cov(\epsilon_E, \epsilon_A)}{\sigma_{\hat{y}}^2 + \sigma_\epsilon^2} = \frac{\rho(\hat{y}_E, \hat{y}_A).\sigma_{\hat{y}}^2 + \rho(\epsilon_E, \epsilon_A).\sigma_\epsilon^2}{\sigma_{\hat{y}}^2 + \sigma_\epsilon^2}, \tag{11}$$

which means:

$$\rho(y_E, y_A) = \rho(\hat{y}_E, \hat{y}_A) \iff \rho(\hat{y}_E, \hat{y}_A) = \rho(\epsilon_E, \epsilon_A). \tag{12}$$

[4]If distributions of earnings are not assumed equal for both generations, the squared of the IGE is just replaced in equation (9) by the squared of the intergenerational correlation (IGC).

So if we assume that the sibling association in observed characteristics used to predict earnings is the same as in non observed characteristics, there is no impact of the use of predicted variables instead of observed ones, on the estimated sibling correlations. We further investigate the difference between sibling correlations in permanent and predicted earnings in Section **??**, with the comparison of French and Swedish estimates.

3 Data

3.1 Description of the database and selection strategy

The data used in this paper come from the French Education-Training-Employment (FQP) survey. Targeted individuals are 18 to 65 year old people living in France, yielding a sample of around 40,000 individuals. We use as our main data set the wave of 2003, in which information on individuals and one of their siblings is available. Interviewed individuals (referred to as "ego") report personal information, notably on their education, occupation and earnings. They are also asked about their family environment and in particular the number of siblings, among whom one is randomly selected (referred to as "alter"). Individuals then also report some basic demographic and socioeconomic information about this sibling, among which education and occupation, but not earnings. The waves of 1970, 1977, 1985 and 1993 are also used as auxiliary data sets to help predicting continuous outcomes.

For our analysis we select individuals (ego) born between 1943 and 1973, which means 30 to 60 years old in 2003, to target individuals out of school but still in the labor market. We only keep individuals who reported information about a sibling. We allow up to 10 years of age difference between the individual (ego) and his/her sibling (alter). Therefore, siblings (alter) can be born between 1933 and 1983 and are 20 to 70 years old in 2003. This choice is made to avoid sampling young people with only older siblings, and old people with only younger siblings.

Available information concerning gender, birth cohort, education level and socio-professional category for both siblings – and earnings for ego – allow us to investigate sibling correlations in different socioeconomic outcomes. Additional information on the composition of the family – as number and birth order of brothers/sisters, age difference between ego and alter – and birth cohort, education, profession of the parents, enable taking various characteristics of the family into account to investigate their impact on sibling correlations.

In order to estimate linear sibling correlations, as seen in Section 2, we need continuous outcomes. In terms of education and occupation, continuous variables of years of education and prestige scores associated with the profession are constructed based on education level and socio-professional category. The predictions of these outcomes are respectively based on OLS regressions and scales of prestige scores. Earnings profiles are estimated to predict annual earnings at age 40 for both siblings using information on education and occupation of both ego and alter, and earnings of ego. OLS regressions as well as Heckman models are computed, to assess the issue of women's labor market participation.

3.2 Education

The available information for both siblings concerning education level is the highest completed certificate or degree. The corresponding variables in 8 groups (1 to 8 from highest degrees to no degree) are already ordinal ones, so that they are used to estimate polychoric correlation coefficients.

In order to predict as a continuous measure the number of years of education of individual i, $EducationYears_i$, for both siblings (this information being available only for ego, not for alter), we implement an OLS regression including as explanatory variables a dummy variable for the gender G_i, a quartic polynomial in age A_i, dummy variables Edu_i^k for the degree category k of individual i (with "no degree" as omitted category), and all corresponding interactions.

$$EducationYears_i = \alpha_1 G_i + \sum_{j=1}^{4} \alpha_{2j} A_i^j + \sum_{j=1}^{4} \alpha_{3j} G_i * A_i^j + \sum_{k=1}^{7} \alpha_{4k} Edu_i^k$$
$$+ \sum_{k=1}^{7} \alpha_{5k} Edu_i^k * G_i + \sum_{j=1}^{4} \sum_{k=1}^{7} \alpha_{6jk} Edu_i^k * A_i^j + \sum_{j=1}^{4} \sum_{k=1}^{7} \alpha_{7jk} Edu_i^k * G_i * A_i^j + u_i. \quad (13)$$

Waves of the survey of 1993 and 2003 are used in order to predict a number of education years for individuals born from 1933 to 1983. Indeed people can only be up to 65 years old in 2003 with the wave of 2003, and we need siblings "alter" aged up to 70 in 2003, which is why we also need the wave of 1993. The graphic representation for this regression is reported in Figure 1 for men (see Figure 3 in appendix for women).[5] This yields a prediction of 12.7 (std err. 2.84) and 13.0 (std err. 2.85) years of education on average for ego and alter respectively.

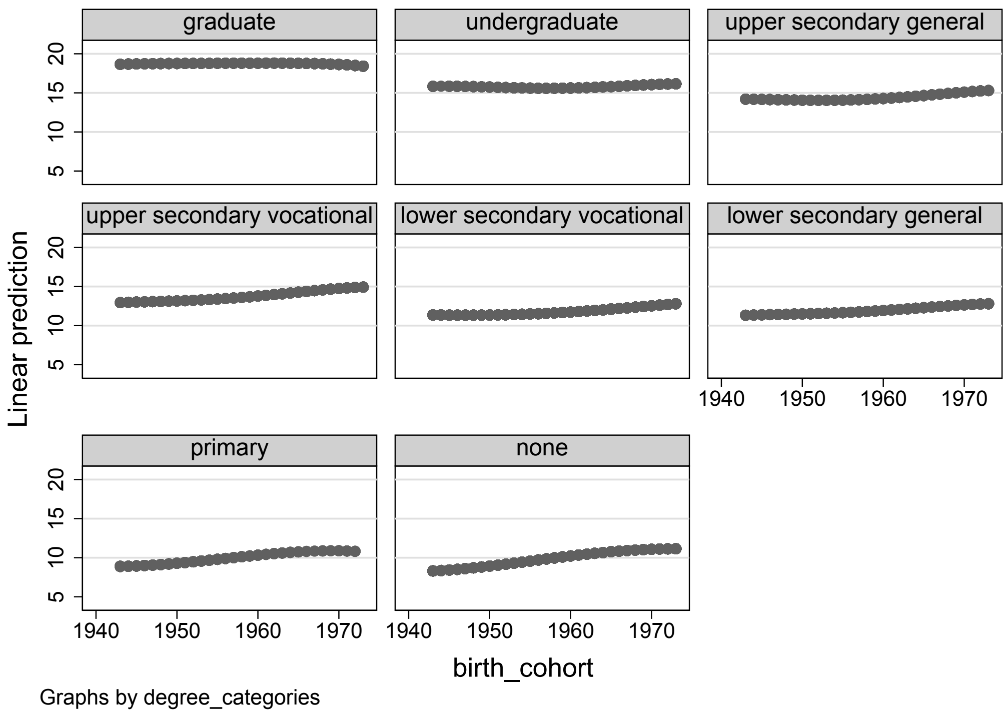


Figure 1: Predicted number of years of education for men

[5]A non-parametric specification including dummies for each gender/cohort/degree category is also tested and yields very similar results.

3.3 Occupation

Regarding occupation, a detailed classification containing 30 socio-professional groups is available in the data and used here. To construct an ordinal variable for the estimation of polychoric correlations, we gather some groups to obtain the following classification: 1) executive, manager, intellectual worker; 2) intermediate occupations; 3) skilled workman, craftspersons, storekeeper, company manager; 4) administrative, sales or service occupations and 5) farmers and laborers.

We also need a continuous measure of the occupation, and thus use prestige scores associated with professions, based on Chambaz et al. (1998). In their paper, they construct scales of prestige scores for different classifications of professions or socio-professional categories. We want to obtain a precise scale by attributing a score to each of our 30 groups. To do so we use the extremely detailed scale of scores attributed to a list of professions. The profession is however only available for ego in our data, so that we attribute the weighted mean of scores (weighted by the frequency of each profession in the groups) for each of our 30 groups of socio-professional categories, for both siblings. The distribution for this classification is reported in Table 12, in appendix.

3.4 Earnings

There is a measure of annual earnings in the wave 2003 of the survey, however only available for interviewed individuals (ego), not for their siblings (alter). The strategy to predict earnings for both siblings is here to estimate earnings profiles in a first step using all observable characteristics that are common to both siblings. We use all waves from 1970 to 2003 (1970, 1977, 1985, 1993 and 2003). Then in a second step we predict log of earnings for both siblings in the database of 2003.

We estimate earnings profiles based on individuals born between years 1933 and 1983 and observed from ages 25 to 55. We normalize age to zero at age 40, at which we predict earnings, in order to avoid life-cycle bias. We construct five groups of birth cohort covering 10 years each (1933-42, 1943-52, 1953-62, 1963-72 and 1973-83) as explanatory variables. The last two groups are actually reunited in the OLS estimation, because the last one contains individuals born from 1973, too late to predict a satisfactory earnings profile, and stops in fact in year 1978, no individual being younger than 25.

The dummy variables Edu_i^k corresponding to the different degree categories of education used in the construction of the continuous number of years of education are also used here as explanatory variables in the prediction of earnings (with again "no degree" as omitted category). Additionally the interactions of education levels with the dummy variables representing cohort groups Coh_i^l (with "1953-62" as omitted category), and with a quadratic function in age A_i, are included as regressors.

The ordinal classification of occupations is used as dummy variables Occ_i^m for interactions with cohort groups (with "laborers" as omitted category). We exclude categories "unknown","farmers" and "skilled workman, craftspersons, storekeeper, company manager", because most individuals of the two last ones are not salaried and therefore do not present a satisfactory measure of earnings. We thus restrict the sample to salaried individuals. We also use a more detailed classification of socio-professional categories as principal effects. This classification contains 16 categories used as dummy variables SPC_i^n (with "unskilled workers" as omitted category) of salaried individuals (only the clerical occupations are additionally excluded).

For men, the regression equation of the log of earnings y_{it} thus contains different age-earnings profiles based on education, as well as interactions between cohort groups and both education and occupation, and can be written:

$$y_{it} = \alpha_t + \sum_{j=1}^{2} \alpha_{1j} A_i^j + \sum_{l=1}^{3} \alpha_{2l} Coh_i^l + \sum_{n=1}^{15} \alpha_{3n} SPC_i^n + \sum_{m=1}^{3} \sum_{l=1}^{3} \alpha_{4ml} Occ_i^m * Coh_i^l$$
$$+ \sum_{k=1}^{7} \alpha_{5k} Edu_i^k + \sum_{k=1}^{7} \sum_{l=1}^{3} \alpha_{6kl} Edu_i^k * Coh_i^l + \sum_{k=1}^{7} \sum_{j=1}^{2} \alpha_{7jk} Edu_i^k * A_i^j + u_{it}, \quad (14)$$

where i and t are indices for individual and date of the survey. Then, we compute predicted log of earnings at age 40 in 2003 for both siblings.

To predict earnings for women, two alternative models are implemented: the same OLS regression as well as an Heckman model, in order to handle the issue of their participation into the labor force. Number of children and spouse's education level, contained in Z_i, are then additionally used to account for the probability of being in salaried employment, with y_{it} only observed for women when the following selection equation is satisfied:

$$f(age_i, Coh_i, Edu_i, Occ_i, SPC_i, Z_i) + v_{it} > 0. \tag{15}$$

To illustrate these earnings profiles, we represent earnings gains obtained for each level of education for the different birth cohort groups, as well as the effect of age on earnings also for each level of education and for the cohort group of reference, individuals born between 1953 and 1962. This is reported in Figure 2 for men (see Figure 4 in appendix for women).

3.5 Description of the sample

To obtain our final sample, we only keep individuals reporting education and occupation information for both siblings.[6] This results in a sample of 19,589 sibling pairs, among which 4,901 pairs of brothers and 4,732 pairs of sisters. The remaining 9,956 are mixed pairs. Ages and age differences among pairs of siblings are reported in Table 1. Siblings are aged 44 on average, with an average age difference of 4 years. We also report in this Table the average number of siblings in the family, which amounts 2.9. More precisely 5,489 family count two siblings only, 5,182 count three, 3,251 count four, the remaining 5,667 count five siblings or more.

[6]This strategy does not excessively reduce the sample size and allows to work with a more stable sample. From an initial sample of 21,885 pairs of siblings, 329 present missing occupation information for ego or alter and 2,212 regarding education.

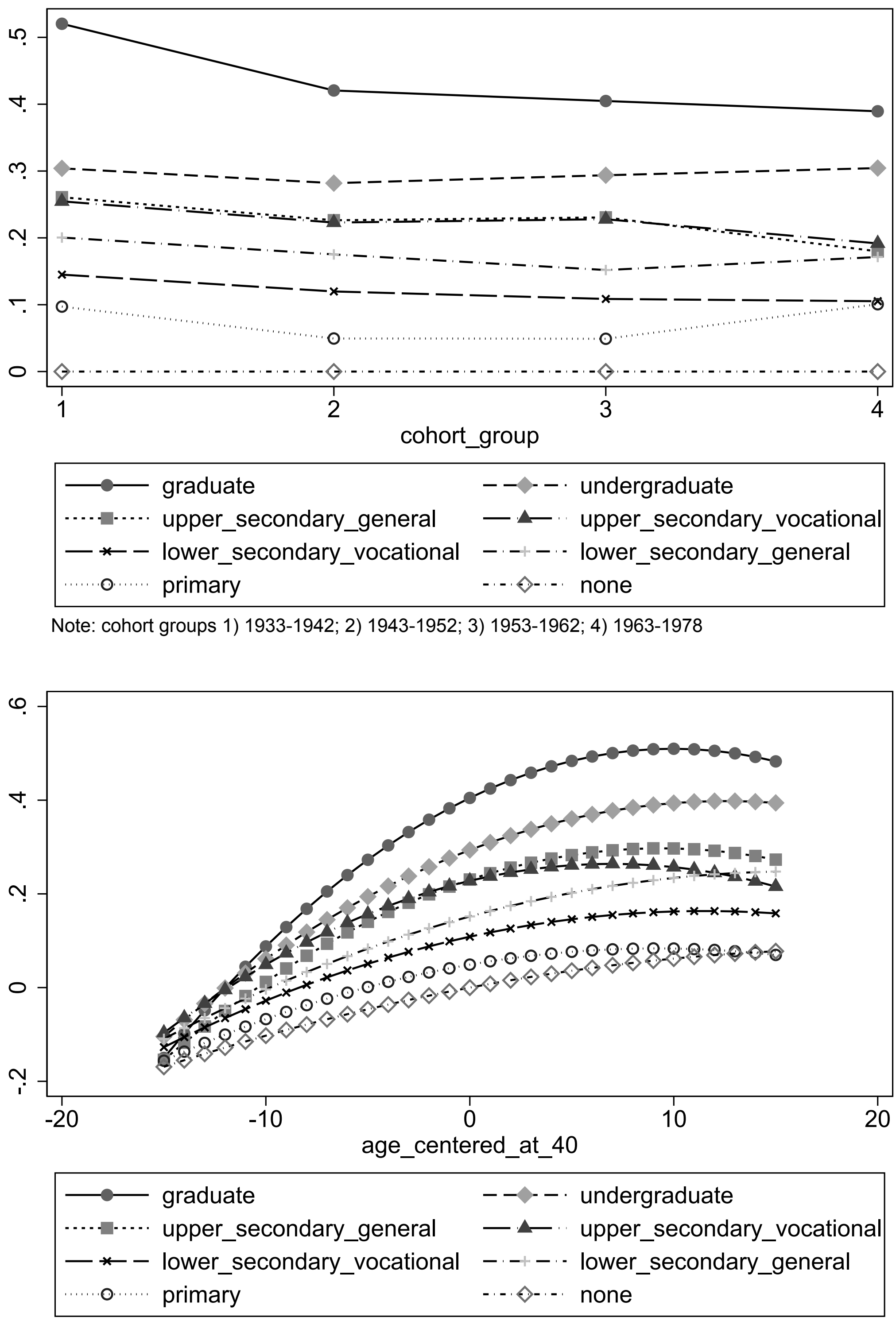


Figure 2: Earnings gains by education and cohort with "no degree" as reference, and returns to age by education for the group reference "born 1953-1962", for men

Table 1: Constitution of the families

	Mean	Std. Dev.	Min	Max	Obs.
Ego's age	44.254	8.603	30	60	19,589
Alter's age	44.267	9.653	20	70	19,589
Age diff.	4.234	2.555	0	10	19,589
0 to 3 years	2.047	0.811	0	3	9,248
4 to 6 years	4.869	0.813	4	6	6,348
7 to 10 years	8.291	1.127	7	10	3,993
Size of sibship	2.941	2.112	1	17	19,589

In Table 2 we present the distributions of the ordinal variables previously described representing degrees and socio-professional categories, for both siblings. For education as well as for occupation, distributions for ego and alter are close.

Table 2: Degrees and socio-professional categories

	Ego		Alter	
	Freq.	Percent	Freq.	Percent
Degree				
graduate	2,196	11.21	2,419	12.35
undergraduate	1,906	9.73	2,382	12.16
upper secondary general	1,302	6.65	1,764	9.01
upper secondary vocational	1,564	7.98	930	4.75
lower secondary vocational	5,411	27.62	6,402	32.68
lower secondary general	1,875	9.57	1,507	7.69
primary	1,682	8.59	1,592	8.13
none	3,653	18.65	2,593	13.24
Category				
executive, ...	2,786	14.22	1,996	10.19
intermediate occupations	4,736	24.18	4,431	22.62
skilled workman, ...	1,024	5.23	1,472	7.51
administrative, ...	6,032	30.79	5,993	30.59
farmer and laborers	5,011	25.58	5,697	29.08

Additionally we also compute ordinal variables representing education and occupation for the parents and the distributions are reported in Table 3. Regarding socio-professional categories, the variables are the same for both generations. For highest completed education, parents are aggregated in only three groups: at least upper secondary, lower secondary, and primary or none.

Table 3: Parental degrees and socio-professional categories

	Father		Mother	
	Freq.	Percent	Freq.	Percent
Degree				
upper secondary or more	2,695	14.13	2,007	10.31
lower secondary	3,739	19.61	2,873	14.75
primary or none	12,635	66.26	14,596	74.94
Category				
executive, ...	1,684	8.87	260	1.96
intermediate occupations	2,655	13.99	1,604	12.11
skilled workman, ...	2,372	12.49	1,215	9.17
administrative, ...	2,084	10.98	5,466	41.25
farmer and laborers	10,189	53.67	4,705	35.51

4 Results

4.1 Main results

We report in Table 4 estimates of polychoric correlations for education and occupation, as well as linear correlations for education, occupation and earnings, both directly using the continuous outcomes (referred to as "gross") and using residuals free of gender and age effects (referred to as "net").[7] The polychoric correlations amount respectively 0.553 and 0.375 for education and occupation, which is close to the gross linear estimates: 0.580 for education and 0.329 for occupation. The estimate is 0.446 for earnings which, when compared to the IGE estimated around 0.5 in Lefranc (2011) – thus corresponding to a sibling correlation of 0.25 if all family influences were accounted for through father's earnings – suggests that a substantial part of the effect of family background was in fact not captured.

These results are satisfactory since it was expected for sibling correlations to be higher in terms of education than occupation. Indeed education is likely to be more affected by family influences as it is determined at an earlier stage in life than occupation. Moreover annual earnings are here predicted based on both education

[7]Results obtained with different strategies to construct the continuous outcomes of education years and annual earnings lead to very similar results: 0.575 (0.005) for the non-parametric approach of predicting education and 0.410 (0.007) with OLS regression used for both gender instead of Heckman model for women, to predict earnings.

and occupation information, therefore it is also not surprising for sibling correlations in terms of earnings to lie in between.[8]

Table 4: Linear and polychoric sibling correlations

	Education		**Occupation**		**Earnings**
	Linear	Polychoric	Linear	Polychoric	Linear
gross	0.580	0.553	0.329	0.375	0.446
	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
net	0.522		0.336		0.459
	(0.005)		(0.006)		(0.006)

Note: Linear corresponds to Pearson's correlations estimated on predicted variables (number of education years, prestige scores and ln of annual earnings at age 40 for education, occupation and earnings respectively), Polychoric corresponds to polychoric correlations estimated on ordinal variables (highest completed certificate or degree and socio-professional categories for education and occupation respectively).

These estimates are also in line with the international literature, as shown in the summary of some recent studies' results in various countries reported in Table 5. In terms of education as well as earnings, our estimates are higher than those of Nordic countries. For education they are smaller than those of the United States and for earnings they are close to those of the United States and Germany. This is coherent with the existing international ranking based on the estimation of intergenerational elasticities.

Whereas controlling for gender and cohort effects only slightly increases sibling correlations in terms of occupation and earnings, from 0.329 to 0.336 and from 0.446 to 0.459 respectively, it decreases the estimates for education from 0.580 to 0.522. An explanation is that education is more affected than occupation or earnings by the fact that siblings are often born in close cohorts, so that a general trend in the evolution of education level artificially raises the sibling correlation. We investigate net correlation coefficients from here on.

[8]The sample is reduced from 19,589 to 16,338 sibling pairs for the estimation of earnings correlations, thus we also estimate education and occupation correlations on this smaller sample: it only slightly increases the net estimates from 0.522 (0.005) to 0.533 (0.006) for education and from 0.336 (0.006) to 0.368 (0.007) for occupation (the evolution being similar on gross estimates).

Table 5: Recent estimates of sibling correlations in education and income

Country	Authors	Data	Cohorts/ages	Est.
Education				
Norway	Björklund and Salvanes (2010)	registers	1962-68	0.40 (0.01)
Sweden	Björklund et al. (2009)	registers	1962-68/30-38	0.48 (0.02)
	Björklund and Jäntti (2012)	registers	1951-67/≈40	0.44 (0.00)
United States	Conley and Glauber (2008)	PSID	1958-76/25-43	0.63 (0.07)
	Mazumder (2008)	NLSY	1957-64/26-41	0.62 (0.01)
	Mazumder (2011)	PSID	1951-68	0.67 (0.03)
Income				
Denmark	Björklund et al. (2002)	registers	1951-68/25-42	0.23 (0.01)
	Schnitzlein (2014)	registers	1952-76/30-50	0.20 (0.01)
Finland	Björklund et al. (2002)	registers	1953-65/25-42	0.26 (0.03)
Germany	Schnitzlein (2014)	SOEP	1952-78/30-50	0.43 (0.08)
Norway	Björklund et al. (2002)	registers	1950-70/25-42	0.14 (0.01)
Sweden	Björklund et al. (2002)	registers	1948-65/25-42	0.25 (0.01)
	Björklund et al. (2009)	registers	1962-68/30-38	0.37 (0.00)
	Björklund and Jäntti (2012)	registers	1951-67/31-40	0.22 (0.00)
United States	Björklund et al. (2002)	PSID	1951-67/25-42	0.43 (0.04)
	Conley and Glauber (2008)	PSID	1958-76/25-43	0.34 (0.07)
	Mazumder (2008)	NLSY	1957-65/26-41	0.49 (0.02)
	Mazumder (2011)	PSID	1951-68	0.51 (0.04)
	Levine and Mazumder (2007)	NLSY	1957-65/26-38	0.45 (0.05)
	Schnitzlein (2014)	PSID	1949-77/30-50	0.45 (0.04)

4.2 Sibling correlations by type of sibling pairs

Different sibling correlations are computed for same-sex (brother/brother and sister/sister) and mixed (brother/sister) sibling pairs and reported in Table 6. As expected, mixed sibling pairs share less than same-sex siblings for each outcome. Sisters seem to have a little more in common than brothers in terms of education and occupation (but these differences are not significant), and less regarding earnings: 0.467 for sisters and 0.517 for brothers.[9]

The relatively low participation of women into the labor force can raise an issue, since prestige scores are attributed according to the last reported socioeconomic category. Mostly for women, this can reflect the professional situation in the beginning of a short career, stopped for instance to raise children, whereas our interest is in obtainable prestige scores, potentially reached if everybody had always worked.

[9] See in appendix the method used for inference issues, based on Fisher (1915).

Table 6: Sibling correlations by gender

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
Brothers	0.543	0.352	0.517
	(0.013)	(0.014)	(0.014)
Sisters	0.551	0.377	0.467
	(0.011)	(0.013)	(0.012)
Mixed pairs	0.497	0.307	0.428
	(0.009)	(0.009)	(0.008)
p-values testing the equality of correlation coefficients			
Brothers/Sisters	0.589	0.159	0.003
Brothers/Mixed	0.000	0.003	0.000
Sisters/Mixed	0.000	0.000	0.010

A first solution to assess this issue is to only take into account currently working women. Therefore, we observe the restricted sample of women (ego) with a brother (alter). Sibling correlations between all women or only working women, and their brothers are reported in Table 7. They are presented for occupation as well as for education, to compare the effects on an outcome potentially affected by the employment of women and the other not. We also compare these results to the same obtained for men (ego) with brothers (alter).[10] As expected for education the results are almost not modified by sampling only currently working individuals as ego. But the differences are also small for occupation. And sampling according to the working status does not change the results more for women than for men.

However a selection problem can rise if the sample is restrained to currently working women. Another method is the investigation of brother/brother-in-law correlations. Again based on the sample of women with a brother, we construct prestige scores for women's spouses (socio-professional categories being available for them too), and we compare them to brothers' ones. The results are also reported in Table 7. The brother-in-law/brother correlation is not very different from, even if slightly lower than sister/brother correlations.

[10]Number of observations for the five groups in Table 7 are respectively 5,525, 4,420, 4,901, 4,527 and 5,525.

Table 7: Sibling correlations for all/working women

	Education	Occupation
All	0.522	0.336
	(0.005)	(0.006)
ego: woman; alter: man	0.494	0.303
	(0.012)	(0.012)
ego: working woman; alter: man	0.483	0.296
	(0.014)	(0.014)
ego: man; alter: man	0.543	0.352
	(0.013)	(0.014)
ego: working man; alter: man	0.551	0.358
	(0.011)	(0.015)
ego: husband; alter: man		0.272
		(0.015)

4.3 Effect of other characteristics on sibling correlations

We also take into account additional parameters, to investigate their impact on sibling correlations. First we want to investigate the evolution over the years of the effect of family background on siblings' outcomes. To do so, we split our sample into three groups, depending on the average parental birth cohort: before 1925, between 1925 and 1935, and after 1935, and estimate different sibling correlations for these three groups. We also test the same strategy based on average siblings' birth cohort: before 1954, between 1954 and 1964, and after 1964.[11] We report in Table 8 the results presenting the evolution of sibling correlations through time, however no clear pattern seems to be observed, so the correlation seems very stable over time.

Family and sibling pair characteristics are then considered, to investigate their effect on sibling correlations (Oettinger, 1999): age difference between ego and alter, number of siblings in the family and whether or not ego or alter is the oldest child of the sibship.[12] Estimates of sibling correlations obtained exploring these factors are reported in Table 9.

[11]Both sets of three groups – constructed based on parental and siblings' birth cohorts respectively – present a nearly perfect repartition in three thirds.

[12]In 12,027 sibling pairs ego or alter is the oldest child of the family, in the 7,562 others it is not the case.

Table 8: Evolution in time of sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
by parental birth cohort			
Before 1925	0.536	0.342	0.479
	(0.010)	(0.011)	(0.011)
1925-1935	0.512	0.346	0.457
	(0.010)	(0.013)	(0.011)
After 1935	0.514	0.321	0.443
	(0.010)	(0.010)	(0.010)
p-values testing the equality of correlation coefficients			
Before 1925/1925-1935	0.059	0.774	0.154
1925-1935/After 1935	0.843	0.096	0.341
Before 1925/After 1935	0.085	0.176	0.016
by siblings' birth cohort			
Before 1954	0.525	0.321	0.467
	(0.012)	(0.012)	(0.011)
1954-1964	0.514	0.351	0.456
	(0.011)	(0.010)	(0.010)
After 1964	0.526	0.333	0.456
	(0.010)	(0.012)	(0.013)
p-values testing the equality of correlation coefficients			
Before 1954/1954-1964	0.412	0.050	0.486
1954-1964/After 1964	0.344	0.255	0.978
Before 1954/After 1964	0.898	0.429	0.476

As expected, age difference has an impact on sibling correlations, at least when comparing closely spaced siblings to those with an important age gap: siblings seem to be more alike when they are about the same age. The estimates fall from 0.541 to 0.471 for education, from 0.347 to 0.312 for occupation and from 0.481 to 0.424 for earnings, for siblings with up to 3 years versus from 7 years age gap.

Concerning the effect of family size, correlations in education and earnings increase with the number of siblings, again the result being significant only when comparing families with substantial different sizes. The correlations increase for instance from 0.471 to 0.529 for education and from 0.410 to 0.440 for earnings, for families counting 2 versus at least 5 siblings.

Lastly the sibling correlation in terms of education is higher, 0.538 versus 0.498,

Table 9: Effect of family and sibling pair characteristics on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
by age difference			
0 to 3 years	0.541	0.347	0.481
	(0.009)	(0.010)	(0.010)
4 to 6 years	0.523	0.334	0.450
	(0.012)	(0.013)	(0.013)
7 to 10 years	0.471	0.312	0.424
	(0.013)	(0.017)	(0.016)
p-values testing the equality of correlation coefficients			
0 to 3/4 to 6	0.123	0.395	0.030
4 to 6/7 to 10	0.001	0.207	0.138
0 to 3/7 to 10	0.000	0.038	0.001
by number of siblings			
2	0.471	0.308	0.410
	(0.012)	(0.014)	(0.015)
3	0.496	0.315	0.437
	(0.013)	(0.012)	(0.013)
4	0.510	0.303	0.438
	(0.013)	(0.015)	(0.014)
5 or more	0.529	0.313	0.440
	(0.013)	(0.013)	(0.013)
p-values testing the equality of correlation coefficients			
2/3	0.092	0.701	0.118
3/4	0.411	0.571	0.967
4/5 or more	0.240	0.627	0.909
2/4	0.021	0.813	0.159
3/5 or more	0.021	0.917	0.858
2/5 or more	0.000	0.774	0.076
whether one is the oldest child			
yes	0.498	0.331	0.441
	(0.008)	(0.009)	(0.008)
no	0.538	0.315	0.457
	(0.009)	(0.010)	(0.011)
p-values testing the equality of correlation coefficients			
yes/no	0.000	0.223	0.228

when neither ego nor alter is the oldest child of the family (Conley, 2009). This would indicate that the oldest child is more different from all other siblings than they are among each other, possibly because he or she is the only one who ever was an only child. The effect of family size may partly be driven by this last result, as it is more likely for either or alter to be the oldest child in smaller families (especially for sibships of only two siblings!).

Finally we want to observe the effect of parental characteristics, such as education and occupation, in order to further assess the impact family background can have on sibling correlations. Thus we report in Tables 10 and 11 the estimated correlation coefficients obtained for each educational level and socio-professional category of both parents. We also estimate these sibling correlations for the whole population, based on residuals net not only from siblings' age and gender effects, but also from education or socio-professional categories of the parents.

We can observe a decrease of sibling correlations in terms of education and earnings, with the increase of educational level of both parents. From lowest to highest completed education of the father, the estimates fall from 0.447 to 0.388 for education and from 0.406 to 0.303 for earnings; from lowest to highest completed education of the mother, they decrease from 0.450 to 0.387 for education and from 0.400 to 0.295 for earnings.

A possible explanation can lie in differences in investment strategies of reinforcement or compensation of sibling differences in initial endowments (Behrman et al., 1982, 1986; Behrman and Taubman, 1986; Behrman et al., 1994) from more or less educated/wealthy parents. Indeed if parents care about the wealth of their children (more than about their earnings), the model of Becker and Tomes (1976) suggests that wealthy parents will invest the most efficient allocation in each child's human capital and then compensate any resulting earnings differences with financial transfers, whereas poorer parents only invest in their children's human capital, taking equality among their children as well as efficiency considerations into account. In this case, sibling differences in human capital and thus earnings are likely to increase with family wealth and education, as we observe.

Concerning the effect of parental occupation, sibling correlations often seem to be lower when parents' socio-professional categories are the highest: *executive, manager, intellectual worker*, which is coherent with the previous interpretation. No other clear pattern is observable.

Table 10: Effect of parental education on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006
Father – net also from father's education	0.419	0.260	0.366
	(0.006)	(0.007)	(0.007)
1) upper secondary or more	0.388	0.248	0.303
	(0.020)	(0.020)	(0.019)
2) lower secondary	0.412	0.232	0.342
	(0.015)	(0.014)	(0.017)
3) primary or none	0.447	0.282	0.406
	(0.007)	(0.008)	(0.008)
p-values testing the equality of correlation coefficients			
1/2	0.271	0.490	0.104
2/3	0.020	0.004	0.000
1/3	0.001	0.083	0.000
Mother – net also from mother's education	0.431	0.267	0.375
	(0.006)	(0.007)	(0.007)
1) upper secondary or more	0.387	0.224	0.295
	(0.023)	(0.021)	(0.021)
2) lower secondary	0.423	0.257	0.351
	(0.017)	(0.016)	(0.016)
3) primary or none	0.450	0.281	0.400
	(0.008)	(0.007)	(0.009)
p-values testing the equality of correlation coefficients			
1/2	0.135	0.223	0.043
2/3	0.103	0.209	0.009
1/3	0.001	0.010	0.000

Table 11: Effect of parental occupation on sibling correlations

	Education	Occupation	Earnings
All	0.522	0.336	0.459
	(0.005)	(0.006)	(0.006)
Father – net also from father's occupation	0.429	0.251	0.360
	(0.006)	(0.007)	(0.007)
1) executive, ...	0.380	0.210	0.303
	(0.020)	(0.022)	(0.023)
2) intermediate occupations	0.447	0.266	0.335
	(0.016)	(0.018)	(0.017)
3) skilled workman, ...	0.452	0.257	0.392
	(0.017)	(0.018)	(0.021)
4) administrative, ...	0.435	0.276	0.384
	(0.018)	(0.019)	(0.018)
5) farmer or laborer	0.428	0.249	0.369
	(0.009)	(0.009)	(0.010)
p-values testing the equality of correlation coefficients			
1/2	0.009	0.055	0.265
2/3	0.840	0.719	0.039
3/4	0.491	0.480	0.783
4/5	0.712	0.221	0.498
1/3	0.006	0.119	0.004
2/4	0.609	0.706	0.069
3/5	0.195	0.718	0.316
1/4	0.042	0.030	0.007
2/5	0.273	0.398	0.096
1/5	0.028	0.115	0.007
Mother – net also from mother's occupation	0.441	0.263	0.376
	(0.007)	(0.008)	(0.009)
1) executive, ...	0.367	0.229	0.229
	(0.057)	(0.058)	(0.057)
2) intermediate occupations	0.445	0.245	0.344
	(0.022)	(0.025)	(0.022)
3) skilled workman, ...	0.428	0.225	0.384
	(0.300)	(0.029)	(0.030)
4) administrative, ...	0.449	0.286	0.389
	(0.011)	(0.012)	(0.013)
5) farmer or laborer	0.440	0.252	0.381
	(0.014)	(0.015)	(0.016)
p-values testing the equality of correlation coefficients			
1/2	0.166	0.801	0.077
2/3	0.577	0.573	0.286
3/4	0.409	0.040	0.855
4/5	0.580	0.068	0.656
1/3	0.296	0.948	0.021
2/4	0.861	0.123	0.081
3/5	0.637	0.367	0.933
1/4	0.125	0.341	0.008
2/5	0.835	0.793	0.175
1/5	0.174	0.702	0.014

5 Conclusion

This paper investigates intergenerational mobility in France through sibling correlations, using data from the French Education-Training-Employment (FQP) survey. We study the impact of family background on different socioeconomic outcomes of adult children – education, occupation and earnings – in order to assess the share of inequalities due to family environment.

First, for two siblings in each family we construct ordinal outcomes of degrees and socio-professional categories, and predict continuous numbers of education years, prestige scores associated with the profession and annual earnings. We then compute polychoric and linear sibling correlations. In the main analysis, we find estimated correlations of 0.522 for education, 0.336 for occupation and 0.459 for earnings.

We also measure the effect of some personal and family characteristics on these sibling correlations. The most significant result is that same-sex sibling pairs share more similarities than mixed pairs. We find that family composition also has an impact. For instance sibling correlations increase with the number of siblings in the family. Finally parental education and socio-professional levels tend to decrease sibling correlations.

Our results allow to compare the situation in France with the recent international literature on sibling correlations. In terms of education, results are a bit higher than 0.4 for Nordic countries, which present a high mobility, and 0.6 for the United States, at the other end of the scale. It is not surprising for our results to lie in between. Concerning earnings, our estimates are close to the German ones. Indeed for Germany sibling correlations in terms of income amount around 0.4 as ours, slightly lower than American ones and higher than the estimates around 0.2 for Nordic countries.

Furthermore our estimated sibling correlations bring a new perspective on the importance of inequality transmission in France, so far investigated with intergenerational elasticities. Indeed the sibling correlation can be expressed as the sum of the squared IGE on the one hand, and the other shared factors uncorrelated to father's earnings on the other hand, as mentioned. Thus if we consider an IGE estimated around 0.5 in Lefranc (2011), which would correspond to a sibling correlation of 0.25 if all family influences were captured through father's earnings, it seems that a large part – around 30% – of the transmission had not been accounted for. So the transmission of inequalities is more important than previously estimated and factors shared by siblings unrelated to parental income play a major role in it.

Thus by presenting sibling correlations for different socioeconomic outcomes, as well as the impact some family characteristics can have on them, this paper constitutes a first step to fill the gap in the literature on sibling correlations in France. It updates the amount and constitution of the French inequality transmission, and confirms the rank of France on this matter between Nordic countries and the United States, and close to other Western European countries.

A Prediction of the outcomes

A.1 Education years

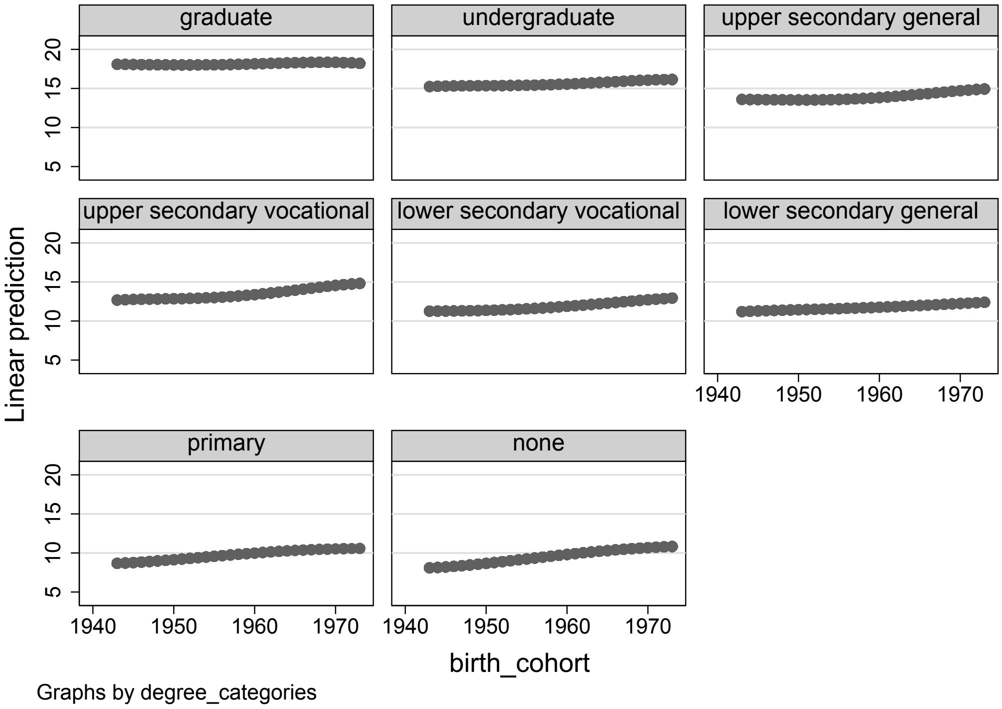


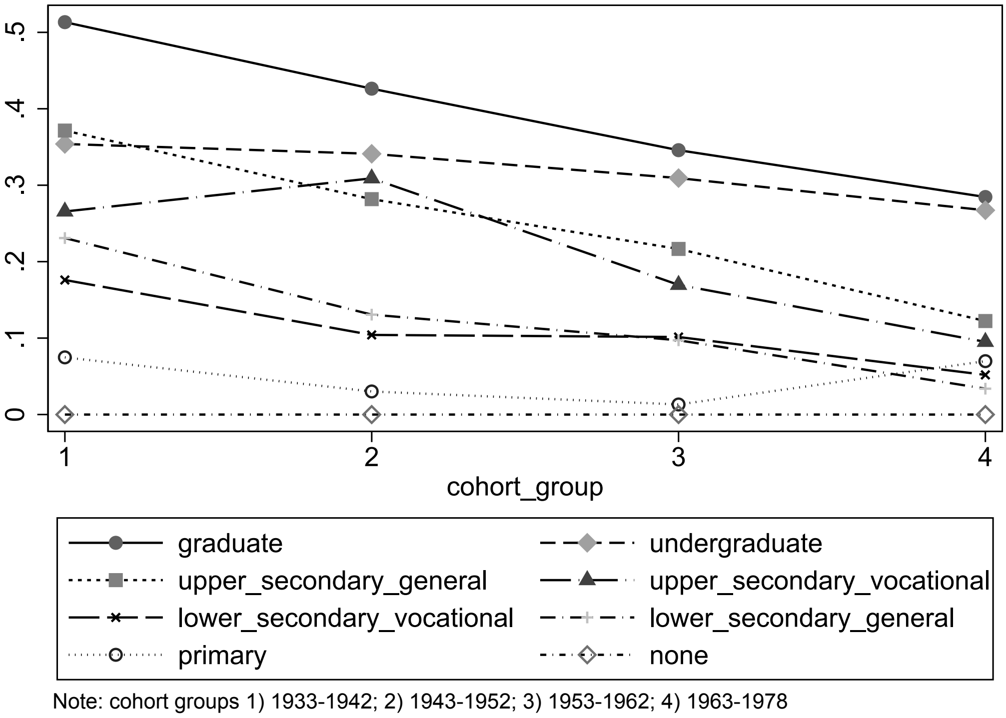
Figure 3: Predicted number of years of education for women

A.2 Prestige scores

Table 12: Prestige score – 30 groups

	Ego		Alter	
Score	Freq.	Percent	Freq.	Percent
-1.694785	566	2.65	163	0.81
-1.563741	1,069	5.00	1,596	7.96
-1.523125	209	0.98	210	1.05
-1.488498	867	4.05	755	3.77
-1.295346	1,520	7.10	1,113	5.55
-0.9188861	2,182	10.20	1,696	8.46
-0.7637425	1,072	5.01	1,514	7.55
-0.7290986	381	1.78	355	1.77
-0.6152064	332	1.55	437	2.18
-0.5838171	1,225	5.73	991	4.94
-0.5739842	533	2.49	479	2.39
-0.3990526	120	0.56	459	2.29
-0.2801967	306	1.43	83	0.41
-0.2024503	154	0.72	32	0.16
-0.1149778	1,464	6.84	1,228	6.12
-0.0760427	1,732	8.09	2,138	10.66
0.0658743	544	2.54	702	3.50
0.138291	485	2.27	668	3.33
0.4168512	643	3.01	390	1.95
0.6803553	219	1.02	223	1.11
0.7463204	838	3.92	661	3.30
0.766371	399	1.86	322	1.61
0.8302992	931	4.35	901	4.49
0.8631468	764	3.57	993	4.95
1.028427	96	0.45	143	0.71
1.296386	298	1.39	247	1.23
1.324646	815	3.81	462	2.30
1.369108	810	3.79	533	2.66
1.40581	619	2.89	282	1.41
1.95731	204	0.95	273	1.36
Total	21,397	100.00	20,049	100.00

A.3 Annual earnings



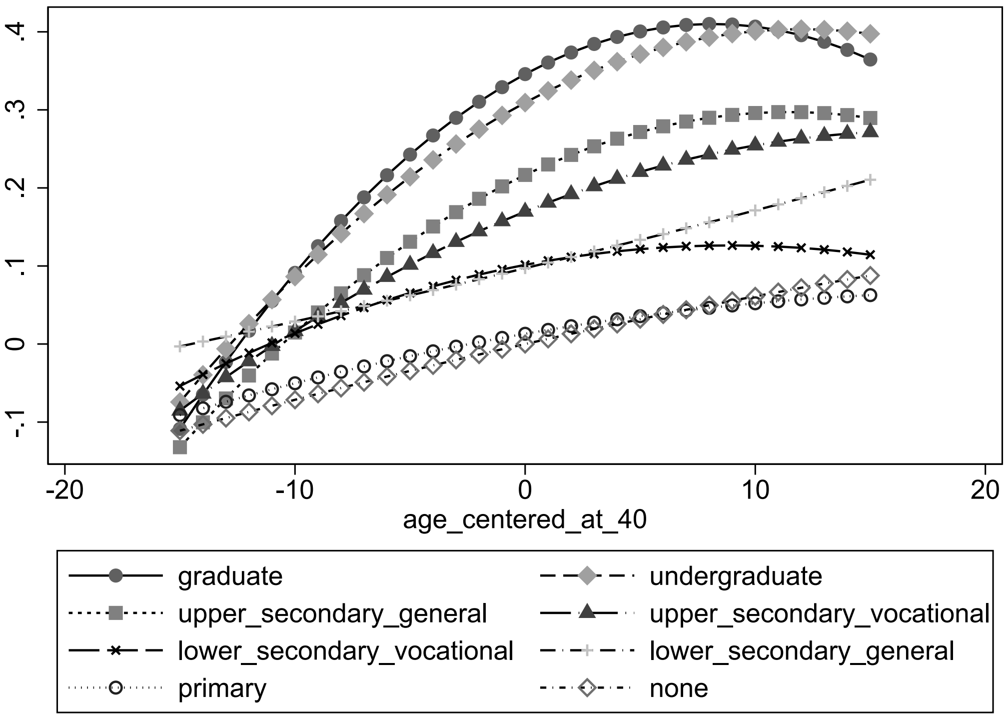


Figure 4: Earnings gains by education and cohort with "no degree" as reference, and returns to age by education for the group reference "born 1953-1962", for women

B Inference in sibling correlations

Pearson's correlation coefficient is approximatively normally distributed for small absolute values of correlation. However for higher values the distribution is skewed. That is why for inference issues we use the so-called Fisher's z transformation to convert Pearson's ρ to the normally distributed variable z, with the standard error σ_z (and number of observations n):

$$z = \frac{1}{2} ln \frac{1+\rho}{1-\rho},$$
$$\sigma_z = \frac{1}{\sqrt{n-3}}.$$

In order to test whether correlation coefficients from two independent groups 1 and 2 are statistically different:

$$H_0 : \rho_1 = \rho_2$$
$$H_1 : \rho_1 \neq \rho_2,$$

we compute the test statistic U, following the standard normal distribution under the null hypothesis:

$$U = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}.$$

References

Altonji, J. and T. Dunn (1991): "Relationships among the family incomes and labor market outcomes of relatives," in *Research in Labor Economics*, ed. by R. Ehrenberg, CT: JAI Press, vol. 12, 269–310.

Becker, G. and N. Tomes (1976): "Child endowments and the quantity and quality of children," *Journal of political Economy*, 84, S143–S162.

Behrman, J., R. Pollak, and P. Taubman (1982): "Parental preferences and provision for progeny," *Journal of Political Economy*, 90, 52–73.

——— (1986): "Do parents favor boys?" *International Economic Review*, 33–54.

Behrman, J., M. Rosenzweig, and P. Taubman (1994): "Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment," *Journal of political economy*, 102, 1131–1174.

Behrman, J. and P. Taubman (1986): "Birth order, schooling, and earnings," *Journal of Labor Economics*, 4, S121–S145.

Bingley, P. and L. Cappellari (2013): "Correlations of brothers earnings and intergenerational transmission," Dipartimento di Economia e Finanza, Università Cattolica.

Björklund, A., T. Eriksson, M. Jäntti, O. Raaum, and E. Österbacka (2002): "Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States," *Journal of Population Economics*, 15, 757–772.

Björklund, A. and M. Jäntti (2009): "Intergenerational income mobility and the role of family background," in *Oxford Handbook of Economic Inequality*, ed. by W. Salverda, B. Nolan, and T. Smeeding, Oxford: Oxford University Press, chap. 20, 491–521.

——— (2012): "How Important is Family Background for Labor-Economic Outcomes?" *Labour Economics*, 19, 465–74.

Björklund, A., M. Jäntti, and M. Lindquist (2009): "Family background and income during the rise of the welfare state: Brother correlations in income for Swedish men born 1932-1968," *Journal of Public Economics*, 93, 671–80.

Björklund, A., L. Lindahl, and M. Lindquist (2010): "What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents," *The BE Journal of Economic Analysis & Policy*, 10, 102.

Björklund, A. and K. Salvanes (2010): "Education and family background: Mechanisms and policies," IZA Discussion Paper.

Black, S. and P. Devereux (2011): "Recent developments in intergenerational mobility," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Vol. 4 Part B, 1487–1541.

Boutchénik, B., C. Coron, S. Grobon, C. Goffette, and L.-A. Vallet (2015): "Quantifying the total influence of family of origin on the schooling and professional careers of individuals," *Economie & Statistique*, 477, 5–23.

Chambaz, C., E. Maurin, and C. Torelli (1998): "L'évaluation sociale des professions en France. Construction et analyse d'une échelle des professions." *Revue française de sociologie*, 39, 177–226.

Conley, D. (2009): *The pecking order: A bold new look at how family and society determine who we become*, Vintage.

Conley, D. and R. Glauber (2008): "All in the family? Family composition, resources, and sibling similarity in socioeconomic status," *Research in Social Stratification and Mobility*, 26, 297–306.

Corcoran, M., R. Gordon, D. Laren, and G. Solon (1990): "Effects of family and community background on economic status," *The American Economic Review*, 80, 362–366.

Corcoran, M., C. Jencks, and M. Olneck (1976): "The effects of family background on earnings," *The American Economic Review*, 66, 430–435.

Drasgow, F. (1988): "Polychoric and polyserial correlations," *Encyclopedia of statistical sciences.*

Fisher, R. (1915): "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, 10, 507–521.

Knigge, A. (2015): *Sources of sibling similarity: Status attainment in the Netherlands during modernization*, ICS-dissertation, Utrecht.

Lefranc, A. (2011): "Educational expansion, earnings compression and changes in intergenerational economic mobility: Evidence from French cohorts, 1931-1976," THEMA Working Papers.

Levine, D. and B. Mazumder (2007): "The growing importance of family: Evidence from brothers' earnings," *Industrial Relations: A Journal of Economy and Society*, 46, 7–21.

Mazumder, B. (2008): "Sibling similarities and economic inequality in the US," *Journal of Population Economics*, 21, 685–701.

——— (2011): "Family and community influences on health and socioeconomic status: sibling correlations over the life course," *The BE Journal of Economic Analysis & Policy*, 11.

Oettinger, G. (1999): "Does the sibling correlation in economic status vary across families and sibling pairs?" Department of Economics, University of Texas.

Schnitzlein, D. (2014): "How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark," *Journal of Population Economics*, 27, 69–89.

Sieben, I. (2001): *Sibling similarities and social stratification: The impact of family background across countries and cohorts*, ICS-dissertation, Nijmegen.

Solon, G. (1999): "Intergenerational mobility in the labor market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Vol. 3 Part A, 1761–1800.

Solon, G., M. Corcoran, R. Gordon, and D. Laren (1991): "A longitudinal analysis of sibling correlations in economic status," *Journal of Human resources*, 509–534.